

Weekly Report

04/20/2015 - 05/03/2015

Jing XIA

May 3, 2015

1 Summary

This week I mainly focus on the DataScanner2 project (the second phase of data inspection).

2 Projects

2.1 Project 1 - DataScanner2

DataScanner2: A system that integrates various data quality metrics, covering all quality metrics categories and providing a quality-aware (clustering, correlation, outliers, complex patterns, image quality, feature preservation) visualization.

Major Reference (Motivation): **Quality Metrics in High-Dimensional Data Visualization: An Overview and Systematization.**

Major features of DataScanner2 are listed below:

Non-orthogonal facet exploration We provide crafted projection (MDS, SVM, etc.) and custom view construction.

Correlation metric The categorization tree is constructed to reveal various relations: mutual information, geometry, cluster, linear trend. Thus we define more relations of a dataset, and maintain the provenance and the explanation of the final categorization tree.

Quality metric DataScanner2 covers various quality-aware metrics (one view quality) from existing data quality metrics (Pixnostics, Scagnostics, Pargnostics, etc): missing data, skewed distribution, inconsistent values, outliers, clusters. It aims at combining automated analysis and visualization techniques for effective exploration of high-dimensional data

Current progress: in implementing Scagnostics, done triangulation, minimal spanning tree, convex hull and alpha hull.

2.2 Project 2 - NBA Game Visualization

-

3 Paper Reading

As listed in project 1.

4 Miscellaneous

1. I passed the driving test.
2. VIS paper review

5 To Do List

1. DataScanner2 — complete Scagnostics implementation.
2. DataScanner2 — start implementing other projects.
3. Group projects plan.